Metabarcoding and Metagenomics 3: 101–115
DOI 10.3897/mbmg.3.39646

Research Article

Assessing pollution of aquatic environments with diatoms' DNA metabarcoding: experience and developments from France Water Framework Directive networks

Vasselon Valentin[1,2,3,4], **Rimet Frédéric**[3], **Domaizon Isabelle**[3], **Monnier Olivier**[5], **Reyjol Yorick**[5,6], **Bouchez Agnès**[3]

1 *Pôle R&D, ECLA, France*
2 *AFB, Site INRA UMR CARRTEL, 75bis avenue de Corzent, 74203 Thonon-les-Bains, France*
3 *INRA, USMB, UMR CARRTEL, 75bis avenue de Corzent, 74203 Thonon-les-Bains, France*
4 *SCIMABIO Interface, 5 rue des Quatre Vents 74200 Thonon-les-Bains*
5 *AFB, 5 Square Félix Nadar, 94300 Vincennes, France*
6 *UMS PatriNat (MNHN-AFB-CNRS), Paris, France*

Corresponding author: *Bouchez Agnès* (agnes.bouchez@inra.fr)

Academic editor: *Dmitry Schigel* | **Received** 2 September 2019 | **Accepted** 29 October 2019 | **Published** 6 November 2019

Abstract

Ecological status assessment of watercourses is based on the calculation of quality indices using pollution sensitivity of targeted biological groups, including diatoms. The determination and quantification of diatom species is generally based on microscopic morphological identification, which requires expertise and is time-consuming and costly. In Europe, this morphological approach is legally imposed by standards and regulatory decrees by the Water Framework Directive (WFD). Over the past decade, a DNA-based molecular biology approach has newly been developed to identify species based on genetic criteria rather than morphological ones (i.e. DNA metabarcoding). In combination with high throughput sequencing technologies, metabarcoding makes it possible both to identify all species present in an environmental sample and to process several hundred samples in parallel. This article presents the results of two recent studies carried out on the WFD networks of rivers of Mayotte (2013–2018) and metropolitan France (2016–2018). These studies aimed at testing the potential application of metabarcoding for biomonitoring in the context of the WFD. We discuss the various methodological developments and optimisations that have been made to make the taxonomic inventories of diatoms produced by metabarcoding more reliable, particularly in terms of species quantification. We present the results of the application of this DNA approach on more than 500 river sites, comparing them with those obtained using the standardised morphological method. Finally, we discuss the potential of metabarcoding for routine application, its limits of application and propose some recommendations for future implementation in WFD.

Key Words

metabarcoding, DNA, High-Throughput Sequencing, diatoms, biomonitoring, ecological status, Water Framework Directive

Introduction

Since it came into force in 2000, the Water Framework Directive (WFD) has provided a common regulatory framework for the implementation of a water management policy in Europe (European Commission 2000). To facilitate their quality assessment, aquatic environments are classified by major water body categories (groundwaters, continental surface waters and coastal waters) and study sites are chosen to set up monitoring networks in each EU member state. This assessment takes into account in particular the ecological status of water-bodies,

by studying the biological characteristics of bioindicator communities living in aquatic environments. In addition to the ecological status, the chemical status of the water is also taken into account to assess the general status of the aquatic environments.

Diatoms are one of the components of phytobenthos, which is a biological quality element that is recommended by WFD for bioassessment of continental surface waters. These microscopic unicellular algae are highly diversified with more than 12,000 species described (Mann and Vanormelingen 2013), each characterised by specific ecological preferences regarding pollution gradients (Pandey et al. 2017). Based on community composition and on relative abundance and ecological preferences of each diatom taxon in the phytobenthic community, it is possible to calculate values for WFD quality indices, such as the Biological Diatom Index (BDI) in France (Coste et al. 2009). These make it possible to highlight pollution of physico-chemical origin, in particular nutrient and organic matter enrichment.

The identification of diatom taxa is traditionally done by optical microscopy through the observation of the morphology of their frustule, a siliceous skeleton that protects the cellular content of each individual. This identification, which requires a high level of expertise, can be time-consuming and costly to achieve the taxonomic resolution required to calculate the indices prescribed by regulation. In France, as the annual assessment of the ecological status of thousands of sites is required for WFD, the use of the morphological approach therefore requires significant skills and resources to carry out this assessment in a robust way. For this reason, new methods for the identification of diatom species, based on DNA techniques, have recently been developed and can both facilitate the ecological assessment and complement the morphological approach (Stein et al. 2014).

Amongst these methods, DNA metabarcoding (Taberlet et al. 2012) potentially identifies all species present in one environmental sample using genetic variability, characterised by a short DNA fragment called a barcode (Hebert et al. 2003; Sogin et al. 2006; Valentini et al. 2009). When DNA metabarcoding is applied on bioindicator communities, the obtained list of environmental DNA sequences can be transformed into a molecular taxonomic inventory, which can then be used to calculate quality indices, similarly to the morphological approach. In addition, the combined use of DNA metabarcoding with high-throughput sequencing (HTS) technologies allows several hundred samples to be processed at the same time, making the molecular approach all the more interesting for bioassessment and monitoring in WFD networks (Keck et al. 2017).

The application of DNA metabarcoding for the characterisation of benthic diatom communities is relatively recent. The first studies, carried out on mock communities, have shown the ability of the molecular approach to produce reliable species inventories (Kermarrec et al. 2013; Zimmermann et al. 2015). Although several subsequent studies have shown the potential of diatom DNA metabarcoding for bioindication, the reliability of the obtained quality scores was still imperfect, which limited its use for assessing the ecological status of watercourses (Kermarrec et al. 2014; Visco et al. 2015). For a given species, the major problem was the lack of a clear correlation between the relative abundance of its DNA sequences obtained in metabarcoding and the relative abundance of its cells counted in microscopy. The main reasons for these quantification discrepancies are summarised in Figure 1. Indeed, exploring a single sample through its morphology or its genetic variability can produce very different pictures of its diatom community due to biological biases (e.g. dead cells or free eDNA, variation in gene copy number or biovolume, cryptic diversity). In addition, there are methodological biases for the metabarcoding approach related, for example, to the DNA extraction method or to bioinformatics processing (Pawlowski et al. 2018).

In the last five years, in order to improve and test the potential of diatom DNA metabarcoding to assess the ecological status of rivers in France, two projects have been carried out on the French WFD monitoring network (Mayotte French overseas department and metropolitan France). The objectives were (i) to identify the biases impacting the molecular results and optimise the DNA workflow, in particular in terms of quantification; (ii) to evaluate on a large scale, more than 500 river sites, the capacity of the molecular approach to produce ecological status assessments comparable to those of the morphological approach.

Here, we present the synthesis of results linked to these 2 projects and originating from several studies (Rimet et al. 2016, 2018, 2019; Vasselon et al. 2017a, b, 2018; Keck et al. 2018; Rivera et al. 2020) and a PhD thesis (Vasselon 2017). This synthesis was first published in the French peer-reviewed journal "Techniques Sciences Méthodes" (Vasselon et al. 2019). We present the methodological developments and choices made to optimise diatom DNA metabarcoding for biomonitoring, in particular the completion of an expert DNA barcode reference library, as well as the optimisation of the quantification. In a second step, we present the results for the application of the optimised molecular approach, at the scale of the biomonitoring of river networks of Mayotte and metropolitan France, by comparing the ecological status assessments provided by the morphological and molecular approaches. Finally, we put into perspective the potential application of the molecular approach for routine surveillance of river WFD networks and resulting issues.

Material and methods

Study site selection

This work is based on two river networks, which were monitored as part of INRA-AFB research projects; one in the French overseas department of Mayotte, the other

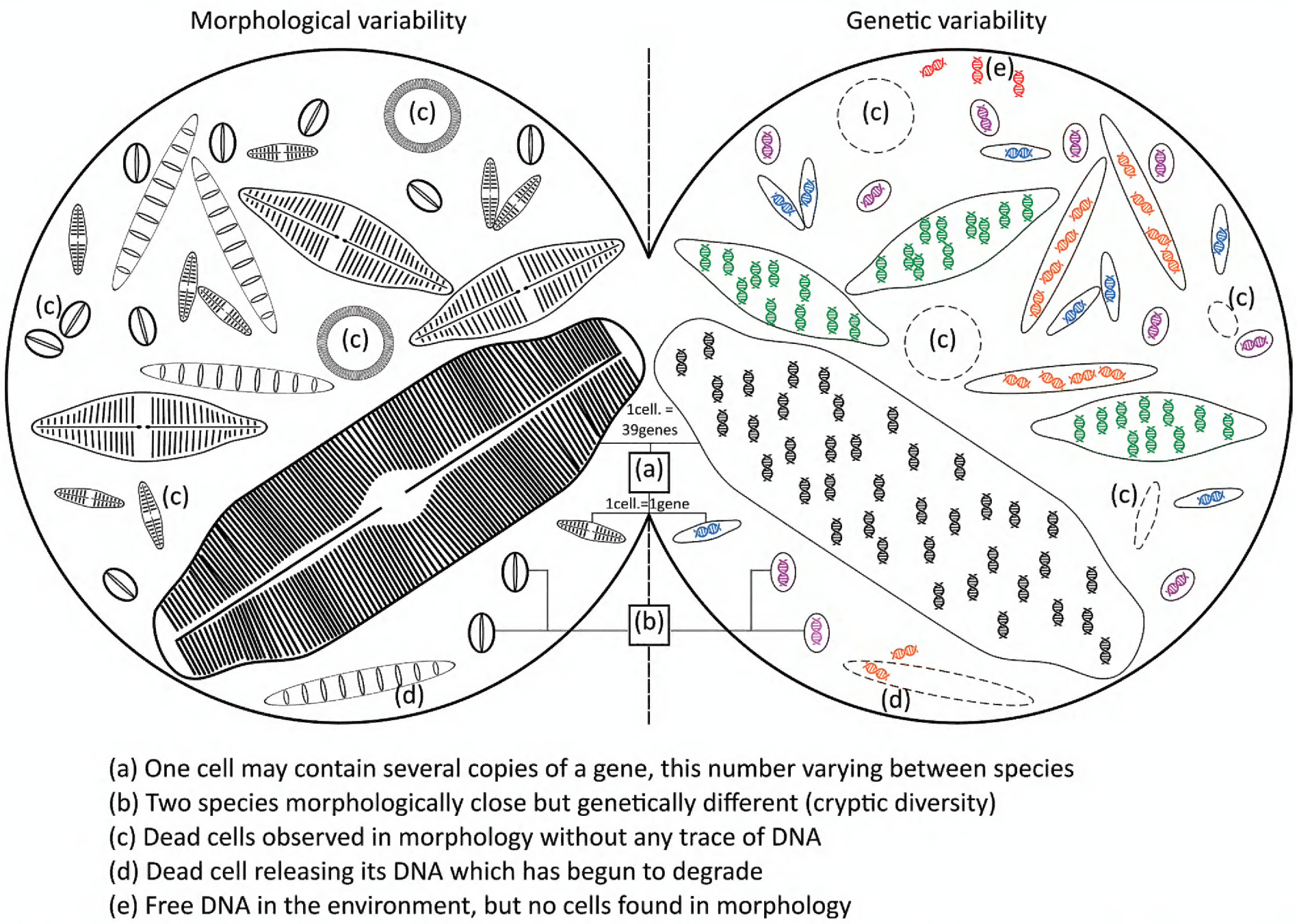

Figure 1. Potential discrepancies between morphological (left) and genetic (right) approaches to characterise the species composition of a single diatom community.

on the French metropolitan territory. These watercourses are all subject to regulatory monitoring under the WFD.

For Mayotte, although the surface area of this French department is modest (376 km²), its rivers present a great diversity of natural and anthropogenic situations and are characterised by various benthic diatom assemblages. A total of 45 stream sites were monitored twice – in years 2014 and 2015 – (Figure 2), the sites representing a marked gradient in water quality, including sites with the lowest possible impact, highly polluted sites and finally intermediate sites belonging to the WFD control network (RCS). Most of the methodological developments of diatom DNA metabarcoding, presented in this article, were performed from this dataset.

For river sites sampled in metropolitan France, they were selected to meet two objectives: (i) complete the DNA barcode reference database for some missing metropolitan important diatom species and (ii) compare ecological status assessments obtained using morphological and molecular approaches on a large scale. In collaboration with the regional environmental services in charge of WFD monitoring (DREAL), 461 sites were chosen to meet these two objectives (Figure 2). For the first purpose (completing the reference database), on the basis of previous morphological inventories, we selected sites known to contain a high abundance of taxa missing in the database (in particular rivers located in the Pyrenees, the Alps, the Massif Central and Brittany). For the second objective, we have selected (i) sites all along the length of the rivers with a marked water quality gradient from upstream to downstream (Seine, Loire, Vienne, Adour, Doubs, Somme, Durance, Allier, Oise, Dordogne and Garonne) and (ii) all the sites of the WFD network in the Rhône catchment area (including the departments of Ain, Jura, Haute-Savoie, Savoie, Rhône and Loire) characterised by a great diversity of environments and pressures (e.g. agricultural, industrial, urban, lowlands and mountainous areas).

Phytobenthos sampling

Aquatic biofilms (periphyton) were collected during the river monitoring campaigns carried out within the framework of the WFD, in agreement with and with the participation of the actors in charge of monitoring (Water Agencies, DREALs, DEALs, consultancies), corresponding to the 2014 and 2015 campaigns for the Mayotte rivers (45 sites, 80 samples) and the 2016 and 2017 campaigns for the metropolitan rivers (461 sites, 461 samples) (Figure 2).

The aquatic biofilm containing the benthic community of diatoms was sampled according to the current standard (NFT 13946, AFNOR 2014). The recommendations of the technical report, CEN/TR 17245 (CEN 2018a) from

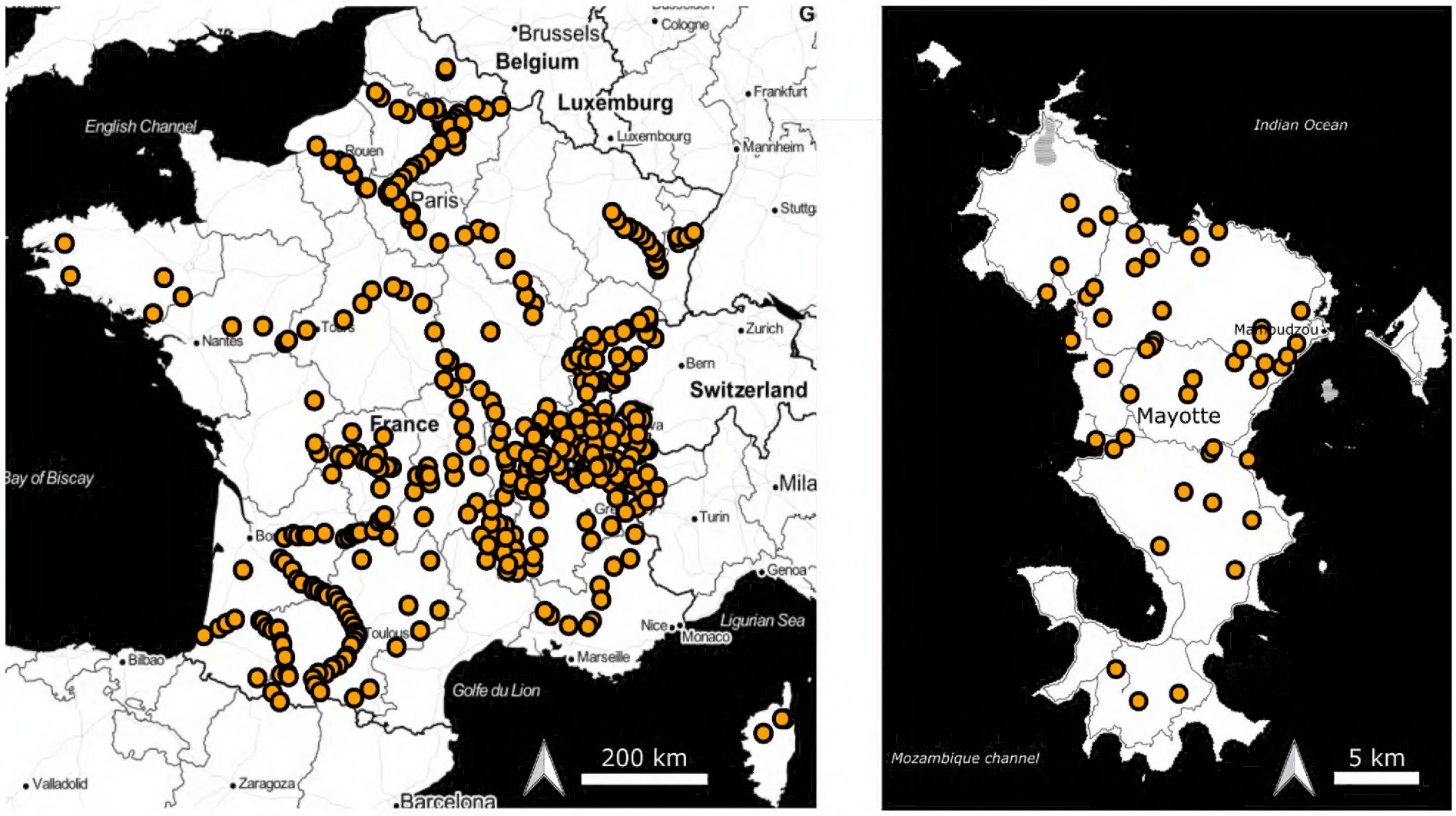

Figure 2. Location of sampled sites on rivers in metropolitan France in 2016 and 2017 (left) and in the overseas department of Mayotte in 2014 and 2015 (right).

the European Committee for Standardisation (CEN) have been followed to ensure that these samples are compatible with the subsequent application of molecular biology techniques. Briefly, the biofilm was brushed from at least five submerged stones located in the stream's lotic zone and fixed in ethanol at a final concentration of at least 70%, which preserves the DNA. In the end, each sample (50 ml) is homogenised and sub-sampled into two batches, for morphological and molecular analysis, respectively.

Morphological approach for diatom inventories

The morphological approach is based on the identification by optical microscopy of diatom taxa based on the morphology of their siliceous external skeleton (frustule). The fixed biofilm samples are prepared in such a way as to allow the determination and counting of diatom species by microscopy, according to the standard NFT 90-354 (AFNOR 2016) (Figure 3). The first step consists of using several successive chemical treatments (H_2O_2, HCl) to remove all organic matter and calcium carbonates and preserve only the diatom siliceous skeleton. Then the treated sample is homogenised and a permanent preparation is carried out by fixing an aliquot of the sample between a slide and a glass lamella in a resin with a high refractive index (Naphrax), thus allowing observation under an optical microscope. Finally, at least 400 diatom valves are identified at species level (or genus when impossible), using floristic guides adapted to Europe for metropolitan samples (e.g. Krammer and Lange-Bertalot 1986; Lange-Bertalot et al. 2017) or tropical regions for Mayotte samples (e.g. Metzeltin and Lange-Bertalot 1998; Tudesque et al. 2008), producing a floristic inventory with the relative abundances of diatom taxa which is used to calculate a quality index for the sampled site.

DNA metabarcoding approach for diatom inventories

The molecular approach in DNA metabarcoding relies on the identification of diatom taxa based on a short DNA fragment (DNA barcode). It requires several successive molecular biology steps to acquire sequence data (DNA extraction, PCR amplification, high-throughput sequencing) and then computer processing of these data to acquire diatom inventories (bioinformatics processing, statistical analyses) (Figure 4). Roughly the same process was used for the samples of the Mayotte and metropolitan rivers; however, some methodological choices differed (protocols, technologies, settings) and are detailed in the scientific articles and synthesis reports corresponding to these two projects (Mayotte: Vasselon et al. 2017b ; Metropolitan France: Keck et al. 2018, Rivera et al. 2020). The most recent and optimised methodology is presented in this article.

DNA extraction is performed from the biofilm pellet obtained after centrifugation of the sample (30 min to 17,000 g) using the NucleoSpin Soil kit (Macherey-Nagel), according to the methodology described by Vasselon et al. (2017b). A 312 bp part of the *rbc*L chloroplastic gene, encoding the large RuBisCo subunit, is used as a DNA barcode with both good specificity and polymorphism for diatoms, following recommendations of Kermarrec et al. (2013). For each sample, at least 2 PCR amplifications of this DNA barcode are performed using a mix of 3 forward primers (Diat_rbcL_708F_1, Diat_rbcL_708F_2, Diat_rbcL_708F_3) and a mix of 2 reverse

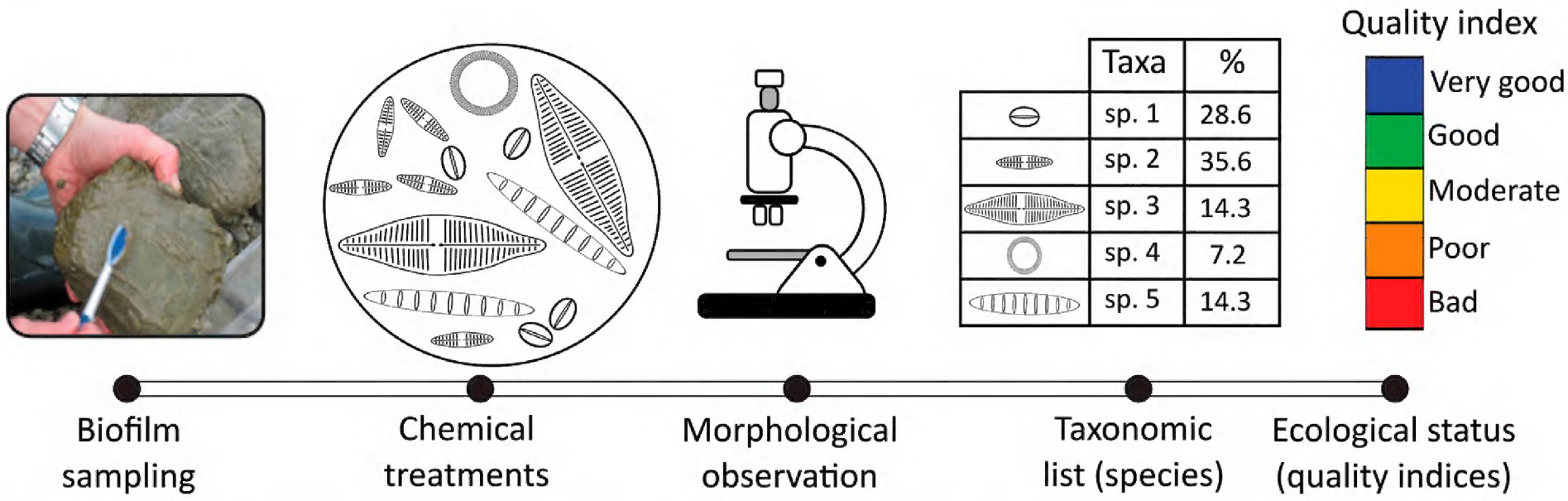

Figure 3. Main steps of the microscopy morphological approach for biomonitoring with diatoms.

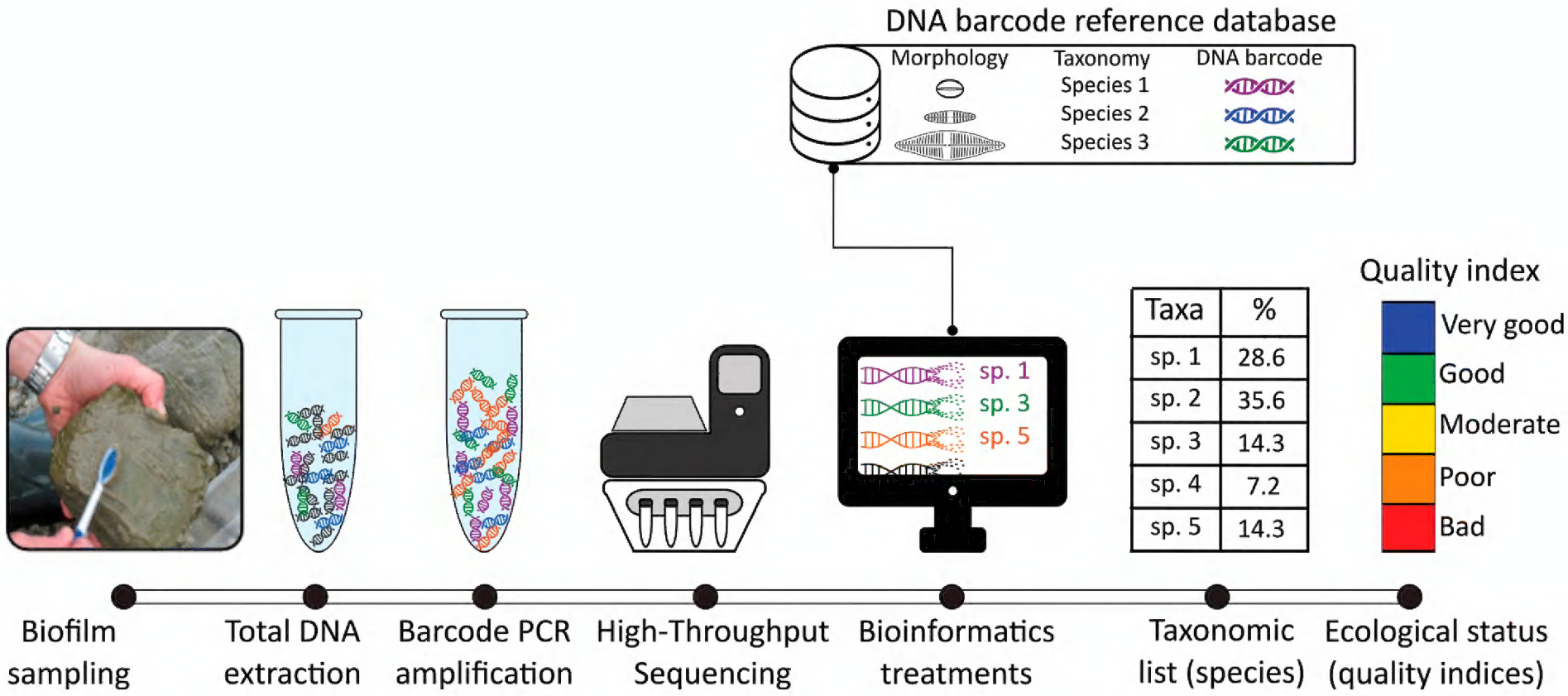

Figure 4. Main steps of the DNA metabarcoding approach for biomonitoring with diatoms.

primers (R3_1, R3_2). The PCR mix and the amplification conditions are described in Keck et al. (2018). The pool of PCR products is then sent to a sequencing platform (e.g. GeT-Plage Toulouse (France), PGTB Bordeaux (France)) which carried out: (i) purification of PCR products; (ii) preparation of sequencing libraries by adding a sample-specific tag and sequencing adapters to the PCR products; (iii) preparation of the final "pool" corresponding to the equimolar mixture of the sample libraries; (iv) paired-end sequencing of the library pool performed with the Illumina Miseq technology (Keck et al. 2018).

The first processing steps of the sequence data ("demultiplexing" and "pair contigating") were performed by the sequencing platform which provides a "fastq" file. All bioinformatics steps were carried out with the Mothur programme (Schloss et al. 2009). The second processing step is dedicated to cleaning sequence data by eliminating sequences of incorrect length, containing errors (e.g. ambiguous bases "N", homopolymers), as well as chimeric sequences. Next, a "de-replication" step allows the grouping of similar sequences together in order to simplify the dataset and facilitate further processing. Only unique sequences represented by more than 2 reads, over the full dataset, are conserved. These de-replicated DNA sequences are then compared with the Diat.barcode reference base (Rimet et al. 2019; previously called RSyst::diatom, Rimet et al. 2016) using a naïve Bayesian classification method (Wang et al. 2007), to assign a taxonomy to each sequence (confidence level > 85%). Only the sequences assigned to the phylum "Bacillariophyta" (diatoms) are kept for further analysis and grouped into OTUs (Operational Taxonomic Units), according to their genetic similarity (> 95% similarity threshold). Finally, a representative DNA sequence and a taxonomy are assigned to each OTU (Keck et al. 2018), providing an inventory of diatom taxa with their relative abundances (estimated by the relative abundances of reads) which is used to calculate a quality index for each sampled site.

Calculation of water quality indices

Although the BDI is the French diatom index used for WFD, the Indice de Polluosensibilité Spécifique (IPS) (Cemagref 1982) was chosen in this study because it includes a greater number of diatom species into its calculation and is frequently used in WFD intercalibration exercises (Coste et al. 2009; Kelly et al. 2014). For each site, morphological and molecular IPS values were calculated

from morphological and molecular inventories using the Omnidia software (version 6.0.2s) (Lecointe et al. 1993).

With regard to molecular inventories, it has been shown that the copy number of the *rbc*L gene is directly correlated to the biovolume of diatom species (Vasselon et al. 2018). Compared to the morphological approach, the molecular approach, therefore, tends to overestimate species with a high biovolume. To limit this bias, a correction factor (CF) was proposed for each species on the basis of its biovolume and applied to correct molecular inventories abundancies in samples from Mayotte and metropolitan France. These corrected taxonomic inventories were then used to calculate a corrected molecular IPS value for each sample.

In the end, three IPS scores were produced, based on the morphological inventory, the molecular inventory, corrected and not corrected by the CF, for each sample.

Results and discussion

Development and optimisation of DNA metabarcoding of diatoms

The use of diatom DNA metabarcoding as a tool for assessing the ecological status of watercourses has been explored over the past 10 years (beginning in France with the PhD thesis of L. Kermarrec (2012)). During this period, complementary studies have made it possible to remove certain methodological obstacles, such as: the validation of the suitability of the *rbc*L DNA barcode for diatoms identification (Kermarrec et al. 2013, 2014), the definition of good practices for the creation of a barcode reference database (Zimmermann et al. 2014) or the implementation of the first complete methodological framework from the biofilm sample to the assessment of the ecological status of the river site from where it was sampled (Kermarrec 2012; Visco et al. 2015; Zimmermann et al. 2015). However, some methodological biases were still limiting the production of reliable molecular taxonomic inventories through DNA metabarcoding, particularly in terms of taxon quantification, thereby producing weak ecological status assessments. As a result, projects were carried out on the French rivers of Mayotte and of the France metropolitan territory in order to consolidate the DNA approach by optimising certain crucial parameters to increase bioassessment efficiency, namely: the coverage of the barcode reference database, the DNA extraction lab step and the quantification of diatom taxa from *rbc*L HTS data.

Development and completion of the barcode reference database "Diat.barcode"

The ability of DNA metabarcoding to accurately identify diatom taxa in an environmental sample is directly related to the quality of the DNA barcode reference database used to assign a taxonomy to OTUs. This relies largely on its diversity coverage, with one quality criterion being its completion (ideal objective of hosting at least one reference DNA barcode per diatom species, Weigand et al. 2019) and another quality criterion being the level of information provided for each barcode (DNA sequences, taxonomic and autecological information, related metadata etc.). Although there are several international reference databases gathering genetic information obtained by the scientific community, such as the National Center for Biotechnology Information (NCBI) (Federhen 2012) or the Barcode Of Life Data system (BOLD) (Ratnasingham and Hebert 2007), the information they contain is generally very heterogeneous in terms of quality (e.g. origin of DNA sequences, wet-lab protocols) and reliability (e.g. traceability of sequences, taxonomic identification).

For this reason, an expert reference database has been developed for diatoms: R-Syst::diatom (Rimet et al. 2016), for which the traditional completion strategy is based on the isolation and culture of environmental cells, their taxonomic identification and the sequencing of the gene of interest from the DNA extracted from the monospecific culture. This diatom database incorporates information about the genes most frequently used as DNA barcodes for diatom metabarcoding: 18S (ribosomal RNA 18S) and *rbc*L. Data from algae culture collections (e.g. Thonon Culture Collection, http://www6.inra.fr/carrtel-collection_eng/), the NCBI database and scientific articles are used to annually update this database. Before being integrated into the database, each new DNA sequence is submitted to several curation steps in order to keep only quality sequences with correct taxonomic identification (Rimet et al. 2016). The data from strains that were isolated, identified and sequenced during the French projects have thus contributed to the development of this database, in particular with the addition of 112 *rbc*L sequences, corresponding to 29 Mayotte diatoms species, some of them tropical or endemic. Despite these efforts, the completeness of this database remains partial as stated by Weigand et al. (2019) in their large gap analysis. For example, of the 100 taxa most frequently identified in morphological inventories during the last monitoring campaigns (1992 to 2014) of French metropolitan rivers, 47 were still absent from the database in 2016 (version 6). As these taxa are very abundant, their weight in the calculation of diatom indices is important and not detecting them is detrimental to the efficiency of molecular indices. To overcome the traditional completion strategy, which is time-consuming and has a moderate success rate, we recently proposed an alternative strategy to recover barcodes from environmental DNA sequences (Rimet et al. 2018). Based on morphological inventories, we focused on low-diversity samples, characterised by diatom assemblages dominated by one taxon absent from the reference database. It was therefore highly likely that a large majority of the metabarcoding DNA sequence reads that were unidentified correspond to this missing taxon. Based on several validation criteria (morphological, genetic, phylogenetic), we associated this environmental DNA

sequence with the taxonomy of the species observed in morphology and integrated it into the reference database (Rimet et al. 2018). This operational approach allowed us to easily and quickly fill crucial gaps in the reference database by targeting abundant taxa, important for index calculation. Thus, from the French metropolitan river samples, we could identify 61 "environmental" *rbc*L barcodes corresponding to 21 species of diatoms, 18 of which were amongst those 100 most frequently identified species in France. Moreover, since 2018, this database has been enriched with collections from other countries (England, Russia) and is managed by a collective of International diatom experts (Rimet et al. 2019), which gives it significant visibility (Pawlowski et al. 2018). This international evolution has led to a change in the name of the reference database now called Diat.barcode, which is hosted and available in open-access on http://www6.inra.fr/carrtel-collection/Barcoding-database.

DNA extraction from benthic diatom communities

The first laboratory step for diatoms metabarcoding is to extract DNA from aquatic biofilm samples. A wide variety of extraction methods and protocols have been developed, depending on the nature of the sample (e.g. water, soil, biofilm, organic tissue) and the targeted biological group (e.g. bacteria, fungi, diatoms) (Dhaliwal 2013). Despite this wide variety of methods, it is sometimes difficult to obtain DNA of good quality and in sufficient quantity to allow the subsequent lab steps (PCR, sequencing) to be carried out. Two main limitations are observed for aquatic biofilms: (i) it is common to co-extract, together with DNA, organic compounds (humic acids), minerals (ions, metals) or enzymes responsible for DNA degradation or PCR inhibition (Schrader et al. 2012); (ii) diatom cells are protected by a silica frustule which may be difficult to break since the strength of the frustule varies from species to species (Hamm et al. 2003; Moreno et al. 2015). It is therefore important to use a method that allows DNA extraction from all diatom species, in order to ensure a good representation of the community.

Five DNA extraction methods frequently used for diatoms were tested on biofilm samples from the Mayotte project (Vasselon et al. 2017a). These methods were compared according to their ability to: (i) extract DNA, regarding both quality and quantity, (ii) obtain molecular inventories using DNA metabarcoding with the *rbc*L barcode that correlates well with taxonomic inventories obtained by microscopy and (iii) provide valid ecological status assessments from diatom indices. Although some disparities could be observed amongst the five methods in terms of quality (presence of PCR inhibitors) and quantity of DNA obtained, none of them prevented the subsequent metabarcoding. All molecular taxonomic inventories provided were identical in terms of specific composition (same diatom taxa detected), but significant variations in the relative abundance of some taxa (e.g. belonging to the genera *Nitzschia*, *Amphora*, *Encyonema* and *Gomphonema*) were observed in two of the five methods. However, these variations had no significant impact on the assessment of the ecological status of the rivers tested, so all the methods tested can be safely used for biomonitoring using diatom metabarcoding (Vasselon et al. 2017a).

Metabarcoding and quantification: impact of the variation in gene copy number

As diatom indices rely on the equation from Zelinka and Marvan (1961), they take into account the relative abundance of taxa. Since the proportions of DNA read sequences obtained in metabarcoding are not clearly correlated to the proportions of cells in microscopy, the ecological status assessments obtained by the two approaches may differ. Amongst the various methodological and biological biases that may potentially influence the relative abundance of DNA read sequences (Pawlowski et al. 2018), the bias related to the variation in the gene copy number appears to have the strongest impact (Figure 1). This bias is dependent on the marker gene used and the targeted biological group, as demonstrated for macroinvertebrates (Elbrecht et al. 2017), fish and amphibians (Evans et al. 2016), oligochaetes (Vivien et al. 2016), foraminifera (Weber and Pawlowski 2013) or microbial communities (Angly et al. 2014). Indeed, Godhe et al. (2008) were able to demonstrate, using quantitative PCR, a correlation between the number of copies of the 18S gene per cell and the cell biovolume of various algae, including diatoms.

We carried out an experiment to verify whether such a correlation existed for the *rbc*L gene and whether it could be used to correct the relative abundances of diatom species in molecular inventories to reconcile them with those obtained by microscopy (Vasselon et al. 2018). Based on tests on monoclonal cultures from 8 diatom species, the results showed that the copy number of the *rbc*L gene varies from one species to another and that this relationship is correlated to cell biovolumes (Figure 5). The equation of the correlation curve could thus be used to make the link between cell biovolume and gene copy number per cell, allowing the calculation of correction factors (CFs) specific to each diatom species. These correction factors were calculated and their efficiency was tested on molecular taxonomic inventories obtained from five mock-communities, made from mixtures of the DNA from the 8 tested species in various controlled proportions. The application of the CFs has made it possible to correct the sequence proportions of each of the 8 species in the molecular inventories, making them comparable to those obtained using the morphological approach.

These CFs were then successfully applied to correct the molecular inventories obtained for the Mayotte rivers WFD network (Vasselon et al. 2018). The calculation method was subsequently refined and new CFs were applied to efficiently correct the molecular inventories from the French metropolitan rivers network (Rivera et al. 2020). The implementation of these quantification corrections is discussed in the following sections of this article.

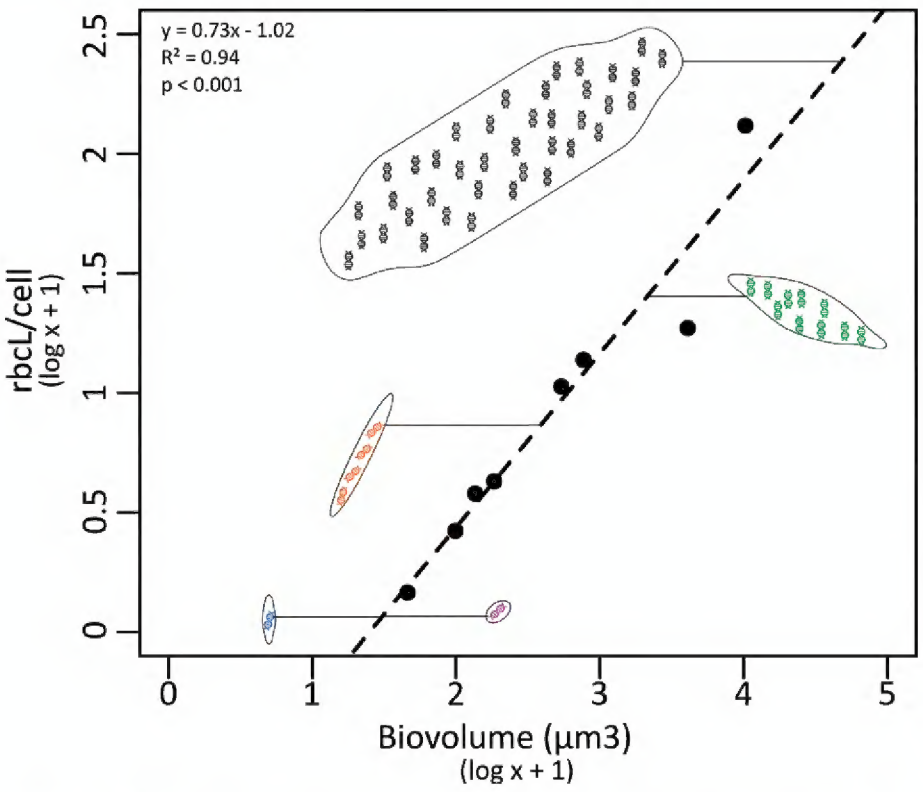

Figure 5. Correlation between the number of copies of the *rbc*L gene per diatom cell and the cell biovolume. Adapted from Vasselon et al. (2018).

Application of the molecular approach to assess the ecological status of watercourses

The DNA metabarcoding optimisations presented above were all integrated at the scale of WFD river networks in Mayotte, France (80 samples) and in metropolitan France (447 samples). These tests were amongst the first to be applied at such a network scale with the objectives of assessing: (i) the large-scale applicability of the molecular approach for diatoms, (ii) its ability to produce taxonomic inventories similar to those obtained with microscopy and (iii) its ability to provide a reliable assessment of the ecological status of rivers.

Potential of diatom DNA metabarcoding to characterise river communities

Molecular inventories have identified 66 diatom species in Mayotte rivers (Vasselon et al. 2017b) and 288 species in those of metropolitan France (Rivera et al. 2020). Morphological taxonomic inventories have identified 204 and 783 species of diatoms in these same rivers, respectively, which is about three times higher. The structures of diatom communities (richness and relative abundances of species), obtained with the two approaches, show a significant but partial correlation (Mantel test; Mayotte: $r = 0.42$, Metropolitan France: $r = 0.63$). There are several reasons for the discrepancies observed between morphological and molecular taxonomic inventories:

Gaps in the reference database: despite the efforts made to complete the DNA barcode reference database, it remains incomplete, avoiding the full assignment of DNA read sequences to diatom species. Unassigned sequences represent 12.4% of the dataset in metropolitan France and 40.7% in Mayotte. This percentage is higher in Mayotte due to the poor knowledge of diatom diversity in this tropical island, which includes endemic and tropical species that are not yet morphologically described (Vasselon et al. 2017b).

Taxonomic differences: the presence of morphologically-related species and the constant evolution of diatom taxonomy (Mann et al. 2016) make it difficult to obtain reliable taxonomic identification with microscopy (Figure 1b). In addition, morphological inventories were produced by several taxonomic experts, which may have led to different identifications of the same morphological entity (species). This phenomenon tends to artificially increase the number of species detected through the morphological approach and thus create variability in ecological status assessments. Therefore, intercalibration exercises between diatom experts are required in order to harmonise assessments (Kahlert et al. 2009).

Detection capacity and limit: Morphological and molecular approaches do not give the same insight into diatom communities and, therefore, do not have the same detection capacity for species (Figure 1a). The relative abundance of small species tends to be overestimated by microscopy (more easily observed in microscopy because they are very abundant), while the molecular approach tends to overestimate the relative abundance of species with a high biovolume (more copies of the *rbc*L gene). Some taxa could therefore not be observed by microscopy, but detected by metabarcoding and vice versa.

False positives: whether through the presence of dead frustules recorded in morphological inventories (Stevenson and Peterson 1991) or the presence of free DNA in the environment from dead individuals (Deiner et al. 2017), different kinds of false positives can be detected by both approaches (Figure 1c–e). Although this artificially increases the discrepancy between methods and the number of species detected, these species generally have a low relative abundance and therefore a moderate impact on final index values.

The taxonomic differences between molecular and morphological inventories mainly affect species present in low abundance. Overall, dominant species, when present in the reference database, are properly detected by both approaches. The application of CFs to molecular inventories has thus made it possible to obtain DNA sequence proportions closer to cell proportions obtained by microscopy. This correction was most effective on molecular inventories from Mayotte samples. This was probably due to the fact that they were strongly dominated by genera with high biovolume (*Eunotia*: 31.9% of sequences, *Ulnaria*: 11.7% of sequences) that were poorly detected in morphological inventories (Vasselon et al. 2017b). The application of CF strongly reduced their importance in molecular inventories, allowing proportions closer to those expected in microscopy to be obtained (reduction in the difference of 99% for *Eunotia* and of 83% for *Ulnaria*).

Potential of diatom DNA metabarcoding to monitor the ecological status of rivers

Despite the compositional differences described previously, the IPS values, calculated from both inventories, were highly correlated (Figure 6). This is because the most dominant species 1) are detected by both approach-

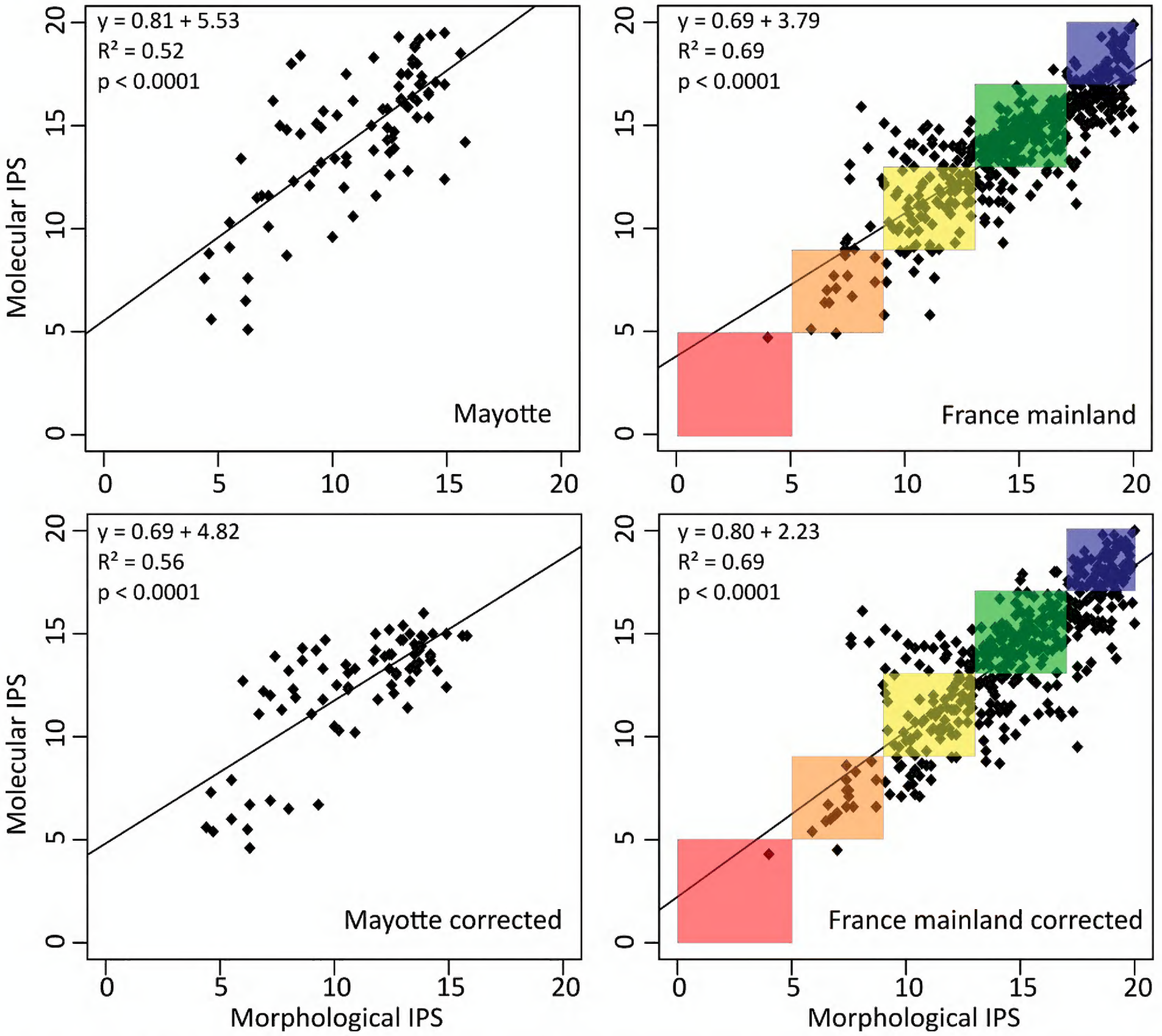

Figure 6. Correlation between morphological and molecular IPS values for samples from the WFD network of Mayotte (left) and of metropolitan France (right). Correction factors (CFs) were (below) or were not (above) applied on molecular taxonomic inventories. The limits of the WFD ecological status classes defined for the IPS are indicated in colour (red = bad, orange = poor, yellow = moderate, green = good, blue = very good), except for Mayotte where these classes are not yet defined. IPS: Indice de Polluosensibilité Spécifique.

es and 2) have the greatest impact on the final index value (Bigler et al. 2009). Other studies on river biofilm samples have also shown this interesting correlation (Visco et al. 2015; Apotheloz-Perret-Gentil et al. 2017; Bailet et al. 2019; Mortágua et al. 2019). In addition, the application of CFs has greatly contributed to reducing the differences in scores between morphological and molecular IPS, improving both the percentage and slope of correlations (Figure 6) (Vasselon et al. 2017b ; Rivera et al. 2020) and has been applied recently on the WFD network of northern Portugal (Mortágua et al. 2019).

Although the IPS index was not developed for the biomonitoring of Mayotte rivers, the corrected morphological and molecular IPS values were both in agreement with the expected quality of the rivers, allowing for a quality gradient ranging from highly impacted to low impacted situations (Figure 6). However, the corrected molecular IPS values were on average 1.9 points higher than the morphological IPS values (Vasselon et al. 2018). This overestimation is partly explained by the lack of some poor quality indicator taxa in molecular inventories. Parts of them were not referenced in the database. Moreover, some of them could still not be identified in metabarcoding, despite the presence of a reference barcode in the database. This is particularly the case of *Nitzschia inconspicua*, which was a paraphyletic species whose taxonomy was poorly defined until recently (Rovira et al. 2015). This taxonomic instability limited its identification in metabarcoding at the genus level. As this species was abundant in some samples (up to 33.5% of the valves observed in microscopy) and is an indicator of rather degraded environments, we failed to correctly identify it, thereby raising the IPS scores. These taxonomic problems were solved in the latest version (version 7) of the Diat.barcode database (Rimet et al. 2019) and this species is now correctly identifiable.

With regard to rivers in metropolitan France, the ecological status obtained using both approaches are congruent (Figure 6). Some 66.2% of the samples belong to the same quality class, 31.8% of the samples have a deviation of one class and only 2% of the samples have a deviation of two classes. In addition, the average difference between the morphological and corrected molecular IPS values is only 1.5. These results are similar to what could be obtained by comparing morphological IPS scores produced by two different diatom experts, with differences of three points already observed for the same sample, corresponding to a difference of one quality class (Kahlert et al. 2012). These results provide additional evidence for the use of DNA metabarcoding as an operational biomonitoring tool to assess the ecological status of watercourses.

Towards a routine use of the molecular approach?

Given all the results from the studies presented here and from other recent tests (e.g. Bailet et al. 2019; Chonova et al. 2019; Mortágua et al. 2019), the use of diatom metabarcoding for ecological assessment seems to be an increasingly promising perspective (Keck et al. 2017; Pawlowski et al. 2018). Recently, England has even abruptly and totally replaced the diatom morphological approach with the molecular one for river monitoring in 2017 (Kelly et al. 2018; Kelly 2019). This still recent and fast developing approach has prompted researchers and biomonitoring stakeholders in Europe to create the COST DNAqua-Net network (Leese et al. 2016, 2018). The objective of this network is to accelerate the process of optimising the molecular approach, harmonise existing methods and protocols between laboratories and define the conditions for implementing molecular approaches within the framework of the WFD. Although the molecular approach gets closer and closer to being technically operational for diatoms, its implementation should guarantee reliable and improved biomonitoring in favour of aquatic ecosystems. This leaves many issues to be addressed with a view to its routine regulatory implementation.

Molecular and morphological approaches are complementary tools

While some recent studies have shown the potential of taxonomy-free approaches (Apotheloz-Perret-Gentil et al. 2017) and machine learning (Tapolczai et al. 2019) to assess water quality using diatom's metabarcoding, one of the questions is to determine the optimum role of molecular tools. These tools will generally come after the morphology-based ones used in the WFD that benefit from a long time series of valuable biomonitoring. In the case of Mayotte rivers, the implementation of the WFD and the development of a diatom index are currently underway (Tapolczai et al. 2017). As the monitoring history in Mayotte is still short, it seems possible to consider using either approach directly, without losing too much history on the evolution of the ecological status of the rivers. In the case of rivers in metropolitan France, monitoring networks have relied on morphological tools for decades (mainly since the 1990s). In such a case, in addition to the political, economic and social consequences that this decision could have (Kelly et al. 2019), it seems difficult to replace it abruptly with the molecular approach without losing valuable information, especially on long-term changes. The loss of time-series continuity could have serious impacts on the actions undertaken over the past several years to recover water quality. Accordingly, it seems preferable, as a first step, to perform both approaches in parallel when long-time series are available.

DNA metabarcoding has the ability to quickly and cost-effectively evaluate a large number of samples. This high-throughput potential could be used, for example, to monitor large numbers of sites as a first step. Then for sites showing higher risk of not achieving good ecological status or for sites subject to management actions to restore good status, it is more crucial to reliably assess their temporal trajectory. In this case, monitoring with the classical approach will ensure continuity in action and in ecological understanding (Kelly et al. 2015), while tiling with the molecular approach will increase the molecular data log and its promises for the future.

Indeed, the molecular approach opens up new perspectives in terms of environmental monitoring. For the time being, the WFD monitoring network for rivers in metropolitan France includes 1,673 sites monitored once a year, for a total river length of 623,464 km, which remains a poor coverage. The molecular approach, with its ability to deal with a large number of samples, offers the perspective to increase: (i) the number of sites monitored annually to provide more spatially complete monitoring and (ii) the number of monitoring operations throughout the year and thus allow the temporal variability of ecosystems to be explored. The latter could be particularly interesting for rivers submitted to large seasonal variations in terms of pressures (e.g. temporary rivers, Stubbington et al. 2019), that could remain undetected with low frequency monitoring, however damaging they may be.

WFD compatibility of molecular approaches for diatoms

In addition to the need to continue the methodological development of the molecular approach, it is also essential to propose strategies to make the ecological status assessments it produces more reliable:

Validate molecular inventories: as with the minimum number of valves required in microscopic counts (AFNOR 2016), it is necessary to define minimum requirements for molecular taxonomic inventories. This can be thresholds, such as a minimum sequencing depth required per sample, taxon detection limits or the maximum proportion of unassigned sequences per sample. The development of such thresholds will increase the reliability and the comparability of molecular inventories.

Control the metabarcoding workflow: a wide variety of methods, technologies and protocols can be used

throughout the workflow of diatom DNA metabarcoding. To validate and compare results, controls should be integrated at each step. For example, it has been proposed to systematically include in each pool of samples, prior to sequencing, a controlled mock-community built from cultures (Kozich et al. 2013). This enables to check whether the whole process has been carried out correctly. Such controls could also make it possible to verify that changes in the metabarcoding process, such as changes in sequencing technology, do not affect the final results.

Intercalibrate and standardise protocols: as for morphological inventories, inter-calibration exercises should be implemented in order to validate methodologies and results from different laboratories and to limit the level of uncertainty in data quality. This should lead to the development of standards which are crucial for the implementation of the molecular approach. For diatom metabarcoding, two CEN technical reports TR 17245 (CEN 2018a) and TR 17244, (CEN 2018b) are already available for the field sampling and the management of reference databases, respectively. The recently launched EcoAlpsWater (INTERREG Alpine Space) project aims at pursuing these efforts at the scale of Alpine region by involving scientists and environmental agencies from 6 countries (France, Italy, Switzerland, Slovenia, Germany and Austria). Further work is required to propose European standards at each step of the workflow, to ensure proper sample preservation, DNA extraction and sequencing, data analysis etc.

Transfer knowledge: it is crucial to set up training courses to enable the transfer of technical and theoretical knowledge to environment managers and stakeholders. The definition and transfer of good practices for implementing the molecular approach will be essential for their effective use in river monitoring and future implementation in environmental regulations.

Deploying the molecular approach in parallel with the morphological during the next WFD monitoring cycle on a large set of WFD river sites will require the definition of a common strategy for: (i) sample collection and storage, (ii) laboratory molecular biology steps (DNA extraction, PCR, sequencing), (iii) bioinformatics processing of DNA sequencing data and (iv) calculation of diatom indices and evaluation of water quality status. Considering the high throughput provided by the molecular approach, it is necessary to meet new needs, such as the sustainability of DNA barcode reference databases, the storage of samples and sequence data and the maintenance and improvement of the molecular approach. To provide a solution for these needs, it is necessary to bring together the different actors (e.g. decision-makers, researchers, managers, monitoring bodies, standardisation bodies, private companies, biotechnology companies), in order to define their respective skills and roles and, thus, to implement this methodology in a concerted and routine manner. This has been initiated at European level within the framework of DNAqua-Net (Hering et al. 2018; Leese et al. 2018) and at a French level thanks to the Mayotte and metropolitan France projects presented in this article. To go further on the implementation of these new monitoring tools, round tables and consultation workshops between actors held in 2019 should provide valuable implementation scenarios (SYNAQUA France-Swiss INTERREG project; Lefrançois et al. 2018). Such efforts should be enlarged to smooth the path towards next generation biomonitoring in Europe.

Conclusion

The rapid and constant technological developments in molecular biology in recent years, particularly in terms of DNA sequencing capacity, have provided increasingly powerful tools to characterise the biodiversity of aquatic ecosystems. They now allow new ways of monitoring aquatic ecosystems, by working at a higher rate on larger spatial and temporal scales. On the other hand, these rapid evolutions leave us limited time to get to grip the new tools, define their application limits and standardise their use. If technological transfers are not well thought out and not well implemented for biomonitoring, the potential gain expected could be lost, replaced by a degradation of our knowledge and ability to monitor and protect aquatic ecosystems. The stability of current bioassessment methods and the experience we now have on their use, although with many limitations, should allow us to take the time to safely acquire new tools in parallel and build tomorrow's biomonitoring.

Acknowledgements

We would like to thank the French Agency for Biodiversity (AFB) for funding the studies on the rivers of Mayotte (2013–2018) and metropolitan France (2016–2018), as well as INRA, which directed this research. We would also like to thank all the actors who participated in this work through sampling and morphological analysis: French DREALs (Aquitaine – D. Sagnet, Auvergne – F. Véry, Bourgogne – V. Peeters, Bretagne – G. Gicquiaud, Centre – S. Saadat and C. Karabaghli, Corse – I. Boulier, Franche-Comté – E. Parmentier, Ile-de-France – O. Cortial, Limousin – J.M. Vouters, Lorraine – D. Heudre, Midi-Pyrénées – E. Seigneur, Pays-de-la-Loire – D. Guillard, Rhône-Alpes – R. Chavaux, Normandie – F. Pétel, PACA – V. Vassal, Picardie – D. Fouré, Nord-Pas-de-Calais – N. Zydek), French Water Agencies (Artois-Picardy – C. Lesniak, Rhône-Méditerranée-Corse – L. Imbert and F. Repellini, Adour-Garonne – M. Durand, J.P. Rebillard and M. Saut, Rhin-Meuse – J.L. Matte and G. Demortier, Loire-Bretagne – J. Durocher, Seine-Normandie – M. Berdoulay) and Consultancies (AQUABIO – R. Marcel and B. Fontan, AQUASCOP – J. Vizinet, ECOMA – L. Kermarrec, SAGE – A. Rolland, J.P. Vulliet and C. Geret, GREBE – P. Prompt, Eurofins Environmental Expertise – L. Feret).

We would also like to thank Gilles Gassiole, Kálmán Tapolczai and Sinziana F. Rivera, who made significant contributions to the sampling, acquisition and analysis of morphological and molecular data and we would like to thank COST Action DNAqua-Net (CA15219), supported by the COST (European Cooperation in Science and Technology) programme which supported this work.

We would like to thank ASTEE who rewarded this PhD work and the editors of TSM and MBMG for their enthusiasm to publish it in their respective journals.

References

AFNOR (2014) NF EN 13946: Water quality – Guidance for the routine sampling and preparation of benthic diatoms from rivers and lakes. 1–18.

AFNOR (2016) NFT 90-354: Water quality – Sampling, treatment and analysis of benthic diatoms from streams and canals. 1–63.

Angly FE, Dennis PG, Skarshewski A, Vanwonterghem I, Hugenholtz P, Tyson GW (2014) CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. Microbiome 2: 1–11. https://doi.org/10.1186/2049-2618-2-11

Apotheloz-Perret-Gentil L, Cordonier A, Straub F, Iseli J, Esling P, Pawlowski J (2017) Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. Molecular Ecology Resources 17: 1231–1242. https://doi.org/10.1111/1755-0998.12668

Bailet B, Bouchez A, Franc A, Frigerio JM, Keck F, Karjalainen SM, Rimet F, Schneider S, Kahlert M (2019) Molecular versus morphological data for benthic diatoms biomonitoring in Northern Europe freshwater and consequences for ecological status. Metabarcoding and Metagenomics 3: e34002. https://doi.org/10.3897/mbmg.3.34002

Bigler C, Gälman V, Renberg I (2009) Numerical simulations suggest that counting sums and taxonomic resolution of diatom analyses to determine IPS pollution and ACID acidity indices can be reduced. Journal of Applied Phycology 22: 541–548. https://doi.org/10.1007/s10811-009-9490-1

Chonova T, Kurmayer R, Rimet F, Labanowski J, Vasselon V, Keck F, Illmer P, Bouchez A (2019) Benthic diatom communities in an Alpine river impacted by waste water treatment effluents as revealed using DNA metabarcoding. Frontiers in Microbiology 10: 1–653. https://doi.org/10.3389/fmicb.2019.00653

Cemagref (1982) Étude des méthodes biologiques quantitative d'appréciation de la qualité des eaux. Ministére de l'Agriculture, CEMAGREF, Division Qualité des Eaux, Pêche et Pisciculture: 1–218.

CEN (2018a) CEN/TR 17245: Water quality – Technical report for the routine sampling of benthic diatoms from rivers and lakes adapted for metabarcoding analyses. CEN/TC 230/WG 23 – Aquatic Macrophytes and Algae: 1–8.

CEN (2018b) CEN/TR 17244: Water quality – Technical report for the management of diatom barcodes. CEN/TC 230/WG 23 – Aquatic Macrophytes and Algae: 1–11.

Coste M, Boutry S, Tison-Rosebery J, Delmas F (2009) Improvements of the Biological Diatom Index (BDI): Description and efficiency of the new version (BDI-2006). Ecological Indicators 9: 621–650. https://doi.org/10.1016/j.ecolind.2008.06.003

Deiner K, Bik HM, Mächler E, Seymour M, Lacoursere-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, De Vere N, Pfrender ME, Bernatchez L (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology 26: 5872–5895. https://doi.org/10.1111/mec.14350

Dhaliwal A (2013) DNA Extraction and Purification. https://doi.org/10.13070/mm.en.3.191

Elbrecht V, Peinert B, Leese F (2017) Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. Ecology and Evolution 7: 6918–2926. https://doi.org/10.1002/ece3.3192

European Commission (2000) Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 Establishing a Framework for Community Action in the Field of Water Policy, 1–72.

Evans NT, Olds BP, Renshaw MA, Turner CR, Li Y, Jerde CL, Mahon AR, Pfrender ME, Lamberti GA, Lodge DM (2016) Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. Molecular Ecology Resources 16: 29–41. https://doi.org/10.1111/1755-0998.12433

Federhen S (2012) The NCBI Taxonomy database. Nucleic Acids Research 40: 136–143. https://doi.org/10.1093/nar/gkr1178

Godhe A, Asplund ME, Härnström K, Saravanan V, Tyagi A, Karunasagar I (2008) Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. Applied and Environmental Microbiology 74: 7174–1782. https://doi.org/10.1128/AEM.01298-08

Hamm CE, Merkel R, Springer O, Jurkojc P, Maier C, Prechtel K, Smetacek V (2003) Architecture and material properties of diatom shells provide effective mechanical protection. Nature 421: 841–843. https://doi.org/10.1038/nature01416

Hebert P, Cywinska A, Ball SL, de Waard JR (2003) Biological identifications through DNA barcodes. Proceedings of The Royal Society B 270: 313–321. https://doi.org/10.1098/rspb.2002.2218

Hering D, Borja A, Jones JI, Pont D, Boets P, Bouchez A, Bruce K, Drakare S, Hänfling B, Kahlert M, Leese F, Meissner K, Mergen P, Reyjol Y, Segurado P, Vogler A, Kelly M (2018) Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. Water Research 138: 192–205. https://doi.org/10.1016/j.watres.2018.03.003

Kahlert M, Albert RL, Antilla EL, Bengtsson R, Bigler C, Eskola T, Gälman V, Gottschalk S, Herlitz E, Jarlman A, Kasperoviciene J, Kokocinski M, Luup H, Miettinen J, Paunksnyte I, Piirsoo K, Quintana I, Raunio J, Sandell B, Simola H, undberg I, Vilbaste S, Weckström J (2009) Harmonization is more important than experience – results of the first Nordic-Baltic diatom intercalibration exercise 2007 (stream monitoring). Journal of Applied Phycology 21: 471–482. https://doi.org/10.1007/s10811-008-9394-5

Kahlert M, Kelly M, Albert RL, Almeida SFP, Besta T, Blanco S, Coste M, Denys L, Ector L, Frankova M, Hlubikova D, Ivanov P, Kennedy B, Marvan P, Mertens A, Miettinen J, Picinska-Falttynowicz J, Rosebery J, Tornes E, Vilbaste S, Vogel A (2012) Identification versus counting protocols as sources of uncertainty in diatom-based ecological status assessments. Hydrobiologia 695: 109–124. https://doi.org/10.1007/s10750-012-1115-z

Keck F, Vasselon V, Rimet F, Bouchez A, Kahlert M (2018) Boosting DNA metabarcoding for biomonitoring with phylogenetic estima-

tion of operational taxonomic units' ecological profiles. Molecular Ecology Resources 18: 1299–1309. https://doi.org/10.1111/1755-0998.12919

Keck F, Vasselon V, Tapolczai K, Rimet F, Bouchez A (2017) Freshwater biomonitoring in the Information Age. Frontiers in Ecology and the Environment 15: 266–274. https://doi.org/10.1002/fee.1490

Kelly M, Urbanic G, Acs E, Bennion H, Bertrin V, Burgess A, Denys L, Gottschalk S, Kahlert M, Karjalainen SM, Kennedy B, Kosi G, Marchetto A, Morin S, Picinska-Falttynowicz J, Poikane S, Rosebery J, Schoenfelder I, Schoenfelder J, Varbiro G (2014) Comparing aspirations: intercalibration of ecological status concepts across European lakes for littoral diatoms. Hydrobiologia 734: 125–141. https://doi.org/10.1007/s10750-014-1874-9

Kelly M (2019) Adapting the (fast-moving) world of molecular ecology to the (slow-moving) world of environmental regulation: lessons from the UK diatom metabarcoding exercise. Metabarcoding and Metagenomics (in press).

Kelly MG, Boonham N, Juggins S, Kille P, Mann DG, Pass D, Sapp M, Sato S, Gover R (2018) A DNA Based Diatom Metabarcoding Approach for Water Framework Directive classification of Rivers. Environment Agency, Horizon House, Deanery Road, Bristol, 147 pp.

Kelly MG, Schneider SC, King L (2015) Customs, habits, and traditions: the role of nonscientific factors in the development of ecological assessment methods. Wiley Interdisciplinary Reviews: Water 2: 159–165. https://doi.org/10.1002/wat2.1068

Kermarrec L (2012) Apport des outils de la biologie moléculaire pour l'utilisation des diatomées comme bioindicateurs de la qualité des écosystèmes aquatiques lotiques et pour l'étude de leur taxonomie. PHD Thesis. Université Grenoble Alpes (Grenoble, France).

Kermarrec L, Franc A, Rimet F, Chaumeil P, Frigerio JM, Humbert J, Bouchez A (2014) A next-generation sequencing approach to river biomonitoring using benthic diatoms. Freshwater Science 33: 349–363. https://doi.org/10.1086/675079

Kermarrec L, Franc A, Rimet F, Chaumeil P, Humbert JF, Bouchez A (2013) Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. Molecular Ecology Resources 13: 607–619. https://doi.org/10.1111/1755-0998.12105

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Applied and Environmental Microbiology 79: 5112–5120. https://doi.org/10.1128/AEM.01043-13

Krammer K, Lange-Bertalot H (1986) Bacillariophyceae 1. Teil: Naviculaceae. Süßwasserflora von Mitteleuropa, Bd. 02/5: Bacillariophyceae. Springer Spektrum, 876 pp.

Lange-Bertalot H, Hofmann G, Werum M, Cantonati M (2017). Freshwater benthic diatoms of central Europe: over 800 common species used in ecological assessment. Koltz Botanical Books (Schmitten-Oberreifenderg, Germany), 942 pp.

Lecointe C, Coste M, Prygiel J (1993) Omnidia: software for taxonomy, calculation of diatom indices and inventories management. Hydrobiologia 269: 509–513. https://doi.org/10.1007/BF00028048

Leese F, Altermatt F, Bouchez A, Ekrem T, Hering D, Meissner K, Mergen P, Pawlowski J, Piggott J, Rimet F, Steinke D, Taberlet P, Weigand A, Abarenkov K, Beja P, Bervoets L, Björnsdóttir S, Boets P, Boggero A, Bones A, Borja Á, Bruce K, Bursić V, Carlsson J, Čiampor F, Čiamporová-Zatovičová Z, Coissac E, Costa F, Costache M, Creer S, Csabai Z, Deiner K, DelValls Á, Drakare S, Duarte S, Eleršek T, Fazi S, Fišer C, Flot J, Fonseca V, Fontaneto D, Grabowski M, Graf W, Guðbrandsson J, Hellström M, Hershkovitz Y, Hollingsworth P, Japoshvili B, Jones J, Kahlert M, Kalamujic Stroil B, Kasapidis P, Kelly M, Kelly-Quinn M, Keskin E, Kõljalg U, Ljubešić Z, Maček I, Mächler E, Mahon A, Marečková M, Mejdandzic M, Mircheva G, Montagna M, Moritz C, Mulk V, Naumoski A, Navodaru I, Padisák J, Pálsson S, Panksep K, Penev L, Petrusek A, Pfannkuchen M, Primmer C, Rinkevich B, Rotter A, Schmidt-Kloiber A, Segurado P, Speksnijder A, Stoev P, Strand M, Šulčius S, Sundberg P, Traugott M, Tsigenopoulos C, Turon X, Valentini A, van der Hoorn B, Várbíró G, Vasquez Hadjilyra M, Viguri J, Vitonytė I, Vogler A, Vrålstad T, Wägele W, Wenne R, Winding A, Woodward G, Zegura B, Zimmermann J (2016) DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. Research Ideas and Outcomes 2: 1–24. https://doi.org/10.3897/rio.2.e11321

Leese F, Bouchez A, Abarenkov K, Altermatt F, Borja A, Bruce K, Ekrem T, Čiampor Jr F, Čiamporová-Zaťovičová Z, Costa FO, Duarte S, Elbrecht V, Fontaneto D, Franc A, Geiger MF, Hering D, Kahlert M, Kalamujić Stroil B, Kelly M, Keskin E, Liska I, Mergen P, Meissner K, Pawlowski J, Penev L, Reyjol Y, Rotter A, Steinke D, van der wal B, Vitecek S, Zimmermann J, Weigand AM (2018) Why we need sustainable networks bridging countries, disciplines, cultures and generations for aquatic biomonitoring 2.0: A perspective derived from the DNAqua-Net COST Action. Advances in Ecological Research 58: 63–99. https://doi.org/10.1016/bs.aecr.2018.01.001

Lefrançois E, Apotheloz-Perret-Gentil L, Blancher P, Botreau S, Chardon C, Crepin L, Cordier T, Cordonier A, Domaizon I, Ferrari BJD, Gueguen J, Hustache JC, Jacas L, Jacquet S, Lacroix S, Mazenq AL, Pawlowska A, Perney P, Pawlowski J, Rimet F, Rubin JF, Trevisan D, Vivien R, Bouchez A (2018) Development and implementation of eco-genomic tools for aquatic ecosystem biomonitoring: the SYNAQUA French-Swiss program. Environmental Science and Pollution Research 25: 33858–33866. https://doi.org/10.1007/s11356-018-2172-2

Mann DG, Crawford RM, Round FE (2016) Bacillariophyta. In: Archibald JM, Simpson AGB, Slamovits CH, Margulis L, Melkonian M, Chapman DJ, Corliss JO (Eds) Handbook of the Protists. Springer International Publishing (New York), 62 pp.

Mann DG, Vanormelingen P (2013) An Inordinate Fondness? The Number, Distributions, and Origins of Diatom Species. Journal of Eukaryotic Microbiology 60: 414–420. https://doi.org/10.1111/jeu.12047

Metzeltin D, Lange-Bertalot H (1998) Tropical diatoms of South America I. Iconographia Diatomologica 5: 1–695.

Moreno MD, Ma K, Schoenung J, Davila LP (2015) An integrated approach for probing the structure and mechanical properties of diatoms: Toward engineered nanotemplates. Acta Biomateriali 25: 313–324. https://doi.org/10.1016/j.actbio.2015.07.028

Mortágua A, Vasselon V, Oliveira R, Elias C, Chardon C, Bouchez A, Rimet F, Feio MJ, Almeida SFP (2019) DNA metabarcoding to enhance the bio-assessment of Portuguese rivers using diatoms. Ecological Indicators. https://doi.org/10.1016/j.ecolind.2019.105470

Pandey LK, Bergey EA, Lyu J, Park J, Choi S, Lee H, Depuydt S, Oh YT, Lee SM, Han T (2017) The use of diatoms in ecotoxicology and bioassessment: Insights, advances and challenges. Water Research 118: 39–58. https://doi.org/10.1016/j.watres.2017.01.062

Pawlowski J, Kelly-Quinn M, Altermatt F, Apotheloz-Perret-Gentil L, Beja P, Boggero A, Borja A, Bouchez A, Cordier T, Domaizon I, Feio MJ, Filipe AF, Fornaroli R, Graf W, Herder J, Van Der Hoorn B, Iwan Jones J, Sagova-Mareckova M, Moritz C, Barquin J, Piggott JJ, Pinna M, Rimet F, Rinkevich B, Sousa-Santos C, Specchia V, Trobajo R, Vasselon V, Vitecek S, Zimmermann J, Weigand A, Leese F, Kahlert M (2018) The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. Science of The Total Environment 637–638: 1295–1310. https://doi.org/10.1016/j.scitotenv.2018.05.002

Ratnasingham S, Hebert PDN (2007) BoLD: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7: 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Rimet F, Abarca N, Bouchez A, Kusber WH, Jahn R, Kahlert M, Keck F, Kelly MG, Mann DG, Piuz A, Trobajo R, Tapolczai K, Vasselon V, Zimmermann J (2018) The potential of High-Throughput Sequencing (HTS) of natural samples as a source of primary taxonomic information for reference libraries of diatom barcodes. Fottea 18: 37–54. https://doi.org/10.5507/fot.2017.013

Rimet F, Chaumeil P, Keck F, Kermarrec L, Vasselon V, Kahlert M, Franc A, Bouchez A (2016) R-Syst::diatom: an open-access and curated barcode database for diatoms and freshwater monitoring. Database: baw016. https://doi.org/10.1093/database/baw016

Rimet F, Evgenuy G, Kahlert M, Kelly MG, Kulikovskiy M, Yevhen M, Mann DG, Pfannkuchen M, Vasselon V, Zimmermann J, Bouchez A (2019) Diat.barcode, an open-access curated barcode library for diatoms. Scientific Reports 9(15116):1–12. https://doi.org/10.1038/s41598-019-51500-6

Rivera SF, Vasselon V, Bouchez A, Rimet F (2020) Diatom metabarcoding applied to large scale monitoring networks: optimization of bioinformatics strategies using Mothur software. Ecological Indicators 109: 105775. https://doi.org/10.1016/j.ecolind.2019.105775

Rovira L, Trobajo R, Sato S, Ibanez C, Mann DG (2015) Genetic and Physiological Diversity in the Diatom *Nitzschia inconspicua*. Journal of Eukaryotic Microbiology 62: 815–832. https://doi.org/10.1111/jeu.12240

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75: 7537–7541. https://doi.org/10.1128/AEM.01541-09

Schrader C, Schielke A, Ellerbroek L, Johne R (2012) PCR inhibitors – occurrence, properties and removal. Journal of Applied Microbiology 113: 1014–1026. https://doi.org/10.1111/j.1365-2672.2012.05384.x

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". Proceedings of the National Academy of Sciences 103: 12115–12120. https://doi.org/10.1073/pnas.0605127103

Stein ED, Martinez MC, Stiles S, Miller PE, Zakharov EV (2014) Is DNA Barcoding Actually Cheaper and Faster than Traditional Morphological Methods: Results from a Survey of Freshwater Bioassessment Efforts in the United States? PLoS One 9: e95525. https://doi.org/10.1371/journal.pone.0095525

Stevenson JR, Peterson CG (1991) Emigration and immigration can be important determinants of benthic diatom assemblages in streams. Freshwater Biology 26: 279–294. https://doi.org/10.1111/j.1365-2427.1991.tb01735.x

Stubbington R, Paillex A, England A, Barthès A, Bouchez A, Rimet F, Sánchez-Montoya M, Westwood CG, Datry T (2019) comparison of biotic groups as dry-phase indicators of ecological quality in intermittent rivers and ephemeral streams. Ecological Indicators 97: 165–174. https://doi.org/10.1016/j.ecolind.2018.09.061

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21: 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Tapolczai K, Bouchez A, Stenger-Kovacs C, Padisak J, Rimet F (2017) Taxonomy- or trait-based ecological assessment for tropical rivers? Case study on benthic diatoms in Mayotte island (France, Indian Ocean). Science of The Total Environment 607–608: 1293–1303. https://doi.org/10.1016/j.scitotenv.2017.07.093

Tapolczai K, Keck F, Bouchez A, Rimet F, Kahlert M, Vasselon V (2019) Diatom DNA metabarcoding for biomonitoring: strategies to avoid major taxonomical and bioinformatical biases limiting molecular indices capacities. Frontiers in Ecology and Evolution 7: 1–409. https://doi.org/10.3389/fevo.2019.00409

Tudesque L, Rimet F, Ector L (2008) A new taxon of the section Nitzschiae lanceolatae Grunow: *Nitzschia costei* sp. nov. compared to *N. fonticola* Grunow, *N. macedonica* Hustedt, *N. tropica* Hustedt and related species. Diatom Research 23: 483–501. https://doi.org/10.1080/0269249X.2008.9705771

Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. Trends in Ecology and Evolution 24: 110–117. https://doi.org/10.1016/j.tree.2008.09.011

Vasselon V (2017) Barcoding et bioindication: développement du metabarcoding des diatomées pour l'évaluation de la qualité des cours d'eau. PHD Thesis. Université Grenoble Alpes (Grenoble, France).

Vasselon V, Domaizon I, Rimet F, Kahlert M, Bouchez A (2017a) Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? Freshwater Science 36: 162–177. https://doi.org/10.1086/690649

Vasselon V, Rimet F, Tapolczai K, Bouchez A (2017b) Assessing ecological status with diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte island, France). Ecological Indicators 82: 1–12. https://doi.org/10.1016/j.ecolind.2017.06.024

Vasselon V, Bouchez A, Rimet F, Jacquet S, Trobajo R, Corniquel M, Tapolczai K, Domaizon I (2018) Avoiding quantification bias in metabarcoding: Application of a cell biovolume correction factor in diatom molecular biomonitoring. Methods in Ecology and Evolution 9: 1060–1069. https://doi.org/10.1111/2041-210X.12960

Vasselon V, Rimet F, Domaizon I, Monnier O, Reyjol Y, Bouchez A (2019) Évaluer la pollution des milieux aquatiques avec l'ADN des diatomées: où en sommes-nous ? Techniques Sciences Méthodes 5: 53–70. https://doi.org/10.1051/tsm/201905053

Visco JA, Apotheloz-Perret-Gentil L, Cordonier A, Esling P, Pillet L, Pawlowski J (2015) Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data. Environmental Science and Technology 49: 7597–7605. https://doi.org/10.1021/es506158m

Vivien R, Lejzerowicz F, Pawlowski J (2016) Next-generation sequencing of aquatic oligochaetes: Comparison of experimental commu-

nities. PLoS One 11: e0148644. https://doi.org/10.1371/journal.pone.0148644

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology 73: 5261–5267. https://doi.org/10.1128/AEM.00062-07

Weber A, Pawlowski J (2013) Can abundance of protists be inferred from sequence data: a case study of foraminifera? PLoS One 8: e56739. https://doi.org/10.1371/journal.pone.0056739

Weigand H, Beermann AJ, Čiampor F, Costa FO, Csabai Z, Duarte S, Geiger MF, Grabowski M, Rimet F, Rulik B, Strand M, Szucsich N, Weigand AM, Willassen E, Wyler SA, Bouchez A, Borja A, Čiamporová-Zaťovičová Z, Ferreira S, Dijkstra KDB, Eisendle U, Freyhof J, Gadawski P, Graf W, Haegerbaeumer A, van der Hoorn BB, Japoshvili B, Keresztes L, Keskin E, Leese F, Macher JN, Mamos T, Paz G, Pešić V, Pfannkuchen DM, Pfannkuchen MA, Price BW, Rinkevich B, Teixeira MAL, Várbíró G, Ekrem, T (2019) DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. Science of The Total Environment 678: 499–524. https://doi.org/10.1016/j.scitotenv.2019.04.247

Zelinka M, Marvan P (1961) Zur Prazisierung der biologischen Klassifikation der Reinheit fliessender Gewasser. Archiv fur Hydrobiologie: 57: 389–407.

Zimmermann J, Abarca N, Enke N, Skibbe O, Kusber WH, Jahn R (2014) Taxonomic Reference Libraries for Environmental Barcoding: A Best Practice Example from Diatom Research. PLoS One 9: e108793. https://doi.org/10.1371/journal.pone.0108793

Zimmermann J, Glöckner G, Jahn R, Enke N, Gemeinholzer B (2015) Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. Molecular Ecology Resources 15: 526–542. https://doi.org/10.1111/1755-0998.12336

Data availability

The data from the WFD river network on the French metropolitan territory are available at the following open-access repositories:

- Sampling sites and floristic lists: https://data.inra.fr/dataset.xhtml?persistentId=doi%3A10.15454%2FWNI6FQ
- Raw sequencing data: https://data.inra.fr/dataset.xhtml?persistentId=doi%3A10.15454%2F9EG5Z4

The data from the WFD river network on the French overseas department of Mayotte are available at the following open-access repositories:

- Sampling sites and floristic lists: https://data.inra.fr/dataset.xhtml?persistentId=doi:10.15454/6Z5IAH&version=1
- Raw sequencing data: https://zenodo.org/record/400160#.XbFuKdU6-Uk